



Random forests-based early warning system for bank failures



Katsuyuki Tanaka, Takuji Kinkyo, Shigeyuki Hamori*

Graduate School of Economics, Kobe University, 2-1, Rokkodai, Nada-Ku, Kobe 657-8501, Japan

HIGHLIGHTS

- This paper introduces a random forests-based early warning system (EWS).
- Our approach is significantly more accurate than other conventional EWSs.
- There are 730 banks in danger with assets equivalent to about 95.3 million US dollars in total.

ARTICLE INFO

Article history:

Received 21 June 2016

Accepted 19 September 2016

Available online 5 October 2016

JEL classification:

G1

C5

Keywords:

Random forests

Early warning system

Bank failure

ABSTRACT

This paper introduces a novel random forests-based early warning system for predicting bank failures. We apply this method to the analysis of bank-level financial statements, in order to find patterns that identify banks in danger of failing. The experimental results show that our method outperforms conventional methods in terms of prediction accuracy.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

An early warning system (EWS) that signals a country's vulnerability to financial crises can be particularly useful for developing an effective financial safety net. Following the seminal work of Frankel and Rose (1996) and Kaminsky et al. (1998), numerous studies on EWSs have sought to identify reliable early warning indicators. In particular, the global financial crisis of 2008–09 has re-boosted interest in EWSs, and has stimulated innovative work (Rose and Spiegel, 2011; Frankel and Saravelos, 2012; Shin, 2013).

The standard approach employed in the literature on EWSs is to estimate a multivariate logistic regression, in which the probability of a crisis is related to a set of explanatory variables, such as current account balance, real exchange rates, credit growth, and fiscal balance. However, there are two major deficiencies with this approach. First, explanatory variables must be pre-selected from a wide range of economic indicators based on some prior information. Second, the logistic regression does not readily allow for non-linear or threshold effects of explanatory

variables. Recognizing these limitations of logistic regressions, Ghosh and Ghosh (2002) employ a method of data analysis known as the decision tree. This method uses a sequence of crisis event prediction rules based on a vector of explanatory variables. At each node of the tree, the sample is split into two sub-branches, according to the threshold value of an explanatory variable. The process is repeated until it reaches a terminal node. This method requires no pre-selection of explanatory variables and allows for non-linear effects.

In this paper, we introduce a novel approach that uses random forests to build an EWS. Random forests, which are a variant of decision trees, significantly improve classification accuracy by building a large number of trees instead of a single tree (Breiman, 2001). It circumvents an over-fitting problem by using randomly selected input variables to split each node. Moreover, it performs better with large datasets. Random forests are used in various application areas, including computer vision and bioinformatics. Random forests are popular because they are simple flexible and can be applied to a range of tasks, including classification and regression. Our approach differs from previous works because we use bank-level financial data. While many studies use macroeconomic explanatory variables to build EWSs, recent literature on the global financial crisis indicates that individual banks' financial conditions can explain differences

* Corresponding author.

E-mail address: hamori@econ.kobe-u.ac.jp (S. Hamori).

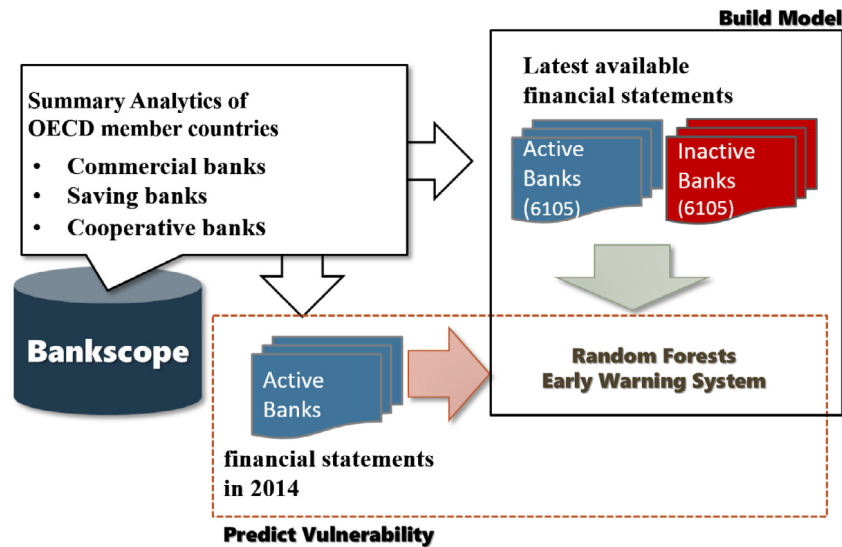


Fig. 1. Overview of random forests EWS.

in cross-bank performance observed during the crises (Berger and Bouwman, 2013; Vazquez and Federico, 2015). To predict bank failure events, we apply the random forests method to the analysis of bank-level financial data in order to identify hidden patterns that can distinguish active and inactive banks. To the best of our knowledge, this is the first paper that employs random forests to build an EWS for predicting bank failure. We call this new type of EWS the Random Forests EWS. We show that the Random Forests EWS outperforms conventional EWSs in terms of prediction accuracy.

The remainder of the paper is organized as follows: Section 2 describes data and methodology. Section 3 evaluates the performance of the Random Forests EWS and demonstrates its usefulness by assessing the vulnerability of OECD member countries based on predicted bank failure. Section 4 presents conclusions.

2. Data and methodology

2.1. Data

The data source for the bank-level financial statements is BankScope. We use 48 indicators derived from the Summary Analytics category. These indicators are classified into four groups: profitability ratio, capitalization, loan quality, and funding. We define a bank failure event as the change of a bank's status from active to inactive (i.e., bankrupt, in liquidation, or dissolved) as reported by BankScope. Our sample covers commercial banks, saving banks, and cooperatives incorporated in OECD Member countries (18,381 in total). We use the latest available financial statements of each bank; the sample period spans from 1986 to 2014.

2.2. The methodology of random forests

We apply random forests to the analysis of bank-level financial statements in order to identify patterns that distinguish between active and inactive banks, thereby producing the prediction for bank failure. More specifically, we use this method to find explanatory variables (and their threshold values) that best split the data into separate classes representing either active or inactive banks. The random forests method is a variant of decision trees with several desirable features, as discussed by Breiman (2001).

First, random forests perform better in terms of classification accuracy by building a large number of trees instead of only a

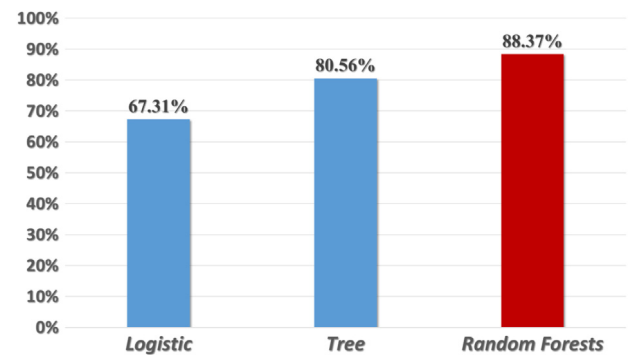


Fig. 2. Average accuracy based on ten-fold cross-validation.

single tree. Each tree is built using randomly selected data samples and randomly selected input variables from the original data to split each node. After a large number of trees are generated, they vote for the most popular class. A single-tree classifier tends to have only slightly better accuracy than a random choice of class. However, by combining a large number of trees using random input selection, the random forests can produce improved accuracy. Second, random forests provide better generalization abilities and are robust against overfitting. This is because using a random selection of input variables to split each node, as well as combining the results of multiple trees, yields error rates that compare favorably to an alternative method and are more robust with respect to noise. Third, random forests perform better with large datasets because multiple trees can be trained efficiently in parallel.

Because the imbalanced size of sample data may provide misleading classification accuracy, we even out the sample sizes of active and inactive banks. As there are fewer inactive banks (6,105 banks), we use only the largest 6,105 active banks in terms of total assets to train our model. Furthermore, we eliminate variables that have missing values more than 50% of the time (there are 34 such variables). Fig. 1 illustrates the overview of the Random Forest EWS.

3. Performance evaluation and prediction

3.1. Evaluation based on ten-fold cross-validation

To evaluate the performance of the Random Forests EWS, we measure the average accuracy of classification using ten-fold cross-

Download English Version:

<https://daneshyari.com/en/article/5058168>

Download Persian Version:

<https://daneshyari.com/article/5058168>

[Daneshyari.com](https://daneshyari.com)